

What is claimed is:

1. A method of training a non-linear regression model of a complex process having operational variables and associated process outcomes, the method comprising the step of:  
determining a connection weight between each of a plurality of output variables and each of a plurality of input variables in the non-linear regression model using a first weight relationship if at least one of the input variable and the output variable comprises a null data value and using a second weight relationship if neither the input variable or the output variable comprise a null data value.

2. The method of claim 1 wherein the determining step comprises:

(a) determining a connection weight between each of a plurality of output variables and each of a plurality of input variables in the non-linear regression model using a first data set that does not comprise a null data value;

(b) determining the connection weight between each of the plurality of output variables and each of the plurality of input variables in the non-linear regression model using (i) a second data that does not comprise a null data value and (ii) the result of step (a); and

(c) determining the connection weight between each of the plurality of output variables and each of the plurality of input variables in the non-linear regression model using (i) a third data set comprising a null data value and (ii) the result of step (b).

3. The method of claim 1 wherein the non-linear regression model comprises at least three layers, each layer having a plurality of nodes, the determining step comprising:

(a) determining a first connection weight between a node of an output layer and a node of a last hidden layer of the non-linear regression model; and

(b) determining a second connection weight between a node of an input layer and a node of a first hidden layer of the non-linear regression model by back-propagating the first connection weight.

4. The method of claim 1, wherein the first weight relationship is of the form:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(w_{ij}(t) - w_{ij}(t-1))$$

where  $w_{ij}(t+1)$  represents a connection weight between a node  $i$  and a node  $j$  for a repetition  $t+1$  of steps (a) and (b),  $w_{ij}(t)$  represents a connection weight between the node  $i$  and the node  $j$  for a

repetition  $t$  of steps (a) and (b) prior to the repetition  $t+1$ ,  $\alpha$  is a momentum coefficient, and  $w_{ij}(t-1)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t-1$  of steps (a) and (b) prior to the repetition  $t$ .

5. The method of claim 4, wherein the momentum coefficient  $\alpha$  is greater than zero and less than or equal to one.

6. The method of claim 1, wherein the second weight relationship is of the form:

$$w_{ij}(t+1) = \eta \frac{\partial E}{\partial w_{ij}(t)} + w_{ij}(t) + \alpha(w_{ij}(t) - w_{ij}(t-1))$$

where  $w_{ij}(t+1)$  represents a connection weight between a node  $i$  and a node  $j$  for a repetition  $t+1$  of steps (a) and (b),  $\eta$  is a learning rate parameter,  $E$  represents an error associated with output of a plurality of nodes  $j$ ,  $w_{ij}(t)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t$  of steps (a) and (b) prior to the repetition  $t+1$ ,  $\alpha$  is a momentum coefficient, and  $w_{ij}(t-1)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t-1$  of steps (a) and (b) prior to the repetition  $t$ .

7. The method of claim 6, wherein the values of the nodes are normalized to have a mean of zero and the learning rate parameter  $\eta$  is greater than zero but less than about 0.5.

8. The method of claim 6, wherein the learning rate parameter  $\eta$  has a value that varies as a function of a number of times a connection weight has been calculated.

9. The method of claim 3 further comprising determining values for a plurality of nodes comprising a gate layer associated with at least one of the at least three layers of the non-linear regression model, each of the plurality of nodes in the gate layer corresponding to one of the plurality of nodes in the associated layer.

10. The method of claim 9 further comprising choosing one of two values for each of the plurality of nodes comprising the gate layer, wherein a first value is chosen if the corresponding node in the associated layer comprises null data and a second value is chosen if the corresponding node in the associated layer does not comprise null data.

11. The method of claim 9 wherein the gate layer is associated with the input layer of the non-linear regression model.
12. The method of claim 9 wherein the gate layer is associated with the output layer of the non-linear regression model.
13. The method of claim 9 wherein the gate layer is associated with a hidden layer of the non-linear regression model.
14. An article of manufacture for training a non-linear regression model of a complex process, the article of manufacture comprising:
  - a process monitor for providing information representing values of a plurality of operational variables and corresponding values of a plurality of process metrics; and
  - a data processing device in signal communication with the process monitor, the data processing device receiving the information and determining a plurality of connection weights to be used in the non-linear regression model from the information,wherein each of the plurality of connection weights is determined using a first weight relationship if the operational variable or corresponding process metric comprises a null data value and using a second weight relationship if neither the operational variable or the corresponding process metric comprise a null data value.
15. The article of manufacture of claim 14, wherein the non-linear regression model comprises at least three layers, each layer having a plurality of nodes,
  - wherein a plurality of nodes of an output layer represents the plurality of process metrics and a plurality of nodes of an input layer represent the plurality of operational variables; and
  - wherein the data processing device determining a first connection weight between a node of an output layer and a node of a last hidden layer of the non-linear regression model from the information, and determining a second connection weight between a node of an input layer and a node of a first hidden layer of the non-linear regression model by back-propagating the first connection weight.
16. The article of manufacture of claim 14 wherein the data processing device further determines if a convergence criterion is satisfied.

17. The article of manufacture of claim 14 wherein each of the plurality of connection weights corresponds to one of the plurality of operational variables and one of the plurality of process metrics.

18. The system of claim 14 wherein the process monitor comprises a database.

19. The system of claim 14 wherein the process monitor comprises a memory device including a plurality of data files, each data file comprising a plurality of scalar numbers representing associated values for nodes of the output layer and the input layer.

20. The system of claim 17 wherein each of the plurality of scalar numbers is normalized with a zero mean.

21. The system of claim 14 wherein first weight relationship implemented by the data processing device is of the form:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(w_{ij}(t) - w_{ij}(t-1))$$

where  $w_{ij}(t+1)$  represents a connection weight between a node  $i$  and a node  $j$  for a repetition  $t+1$  of steps (a) and (b),  $w_{ij}(t)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t$  of steps (a) and (b) prior to the repetition  $t+1$ ,  $\alpha$  is a momentum coefficient, and  $w_{ij}(t-1)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t-1$  of steps (a) and (b) prior to the repetition  $t$ .

22. The system of claim 14 wherein first weight relationship implemented by the data processing device is of the form:

$$w_{ij}(t+1) = \eta \frac{\partial E}{\partial w_{ij}(t)} + w_{ij}(t) + \alpha(w_{ij}(t) - w_{ij}(t-1))$$

where  $w_{ij}(t+1)$  represents a connection weight between a node  $i$  and a node  $j$  for a repetition  $t+1$  of steps (a) and (b),  $\eta$  is a learning rate parameter,  $E$  represents an error associated with output of a plurality of nodes  $j$ ,  $w_{ij}(t)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t$  of steps (a) and (b) prior to the repetition  $t+1$ ,  $\alpha$  is a momentum coefficient, and  $w_{ij}(t-1)$  represents a connection weight between the node  $i$  and the node  $j$  for a repetition  $t-1$  of steps (a) and (b) prior to the repetition  $t$ .